PROGRESS IN DEVELOPMENTAL BIOLOGY
Part A

Proceedings of the Tenth International Congress of the International Society of Developmental Biologists, held in Los Angeles, CA, August 4–9, 1985

Editor

Harold C. Slavkin
Laboratory for Developmental Biology
School of Dentistry
University of Southern California
Los Angeles, California

ALAN R. LISS, INC. • NEW YORK

Address all Inquiries to the Publisher
Alan R. Liss, Inc., 41 East 11th Street, New York, NY 10003

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

International Society of Developmental Biologists.
International Congress (10th : 1985 : Los Angeles, Calif.)
Progress in developmental biology.
(Progress in clinical and biological research ; 217A)
Includes index.
1. Developmental biology—Congresses. I. Slavkin, Harold C. II. Title. III. Series.
QH491.I48 1985 574.3 86-7483
ISBN 0-8451-0192-7 (v. 1)
ISBN 0-8451-5067-7 (set)

Progress in Developmental Biology, Part A, pages 205-218

GEOFFROY AND THE HOMEOBOX

Stephen Jay Gould

Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138

As Dickens said about another moment in French history, July, 1830 was both the best and the worst of times. During July 27-29, known ever since to Frenchmen as Les Trois Glorieuses, citizens manned the barricades in Paris and finally forced the abdication of despotic Charles X, younger brother of Louis XVIII. On July 31, on the balcony of the Hotel de Ville, the aging Marquis de Lafayette, America's friend in previous times of trouble, draped himself in the tricolor banner, embraced the Duc d'Orleans and proclaimed him Louis-Philippe, the new constitutional monarch of France.

In the midst of this upheaval, a noted German observer wrote with brimming enthusiasm about a great Parisian battle: "We find in conflict here two such different ways of thinking...It has often happened before in the history of this antagonism, and the phenomenon must occur again and again...When such different elements are forced into contact, where they touch, every time an explosion must result."

The observer, we assume, must be writing about the age-old strife of monarchists and republicans, then playing its latest version on the streets of Paris. He was not. He described, in these dispatches, an argument then raging in l'Academie des Sciences--"of such great importance," he wrote, "that the history of science may never experience a second example."

The debate, running throughout 1830 in and around the

July revolution, featured France's two greatest zoologists as adversaries: Georges Cuvier and Etienne Geoffroy Saint-Hilaire. The observer was no mean journalist: Johann Wolfgang von Goethe. Geoffroy's son, writing later about the great debate, spoke of Goethe's involvement:

> At the very dawn of events that would overthrow a throne and change the face of France, he immersed himself in this peaceful revolution occurring in the pure realm of ideas.

Goethe was no incidental bystander, roped in on a lark. The great poet was also a fine scientist, a student of organic form and inventor, incidentally, of the word "morphology." He was also a partisan of Geoffroy's side in the great debate.

This famous struggle often portrayed quite wrongly as a debate about evolution (which Cuvier opposed to the core and Geoffroy supported in a fitful way), began as an argument about anatomy and ended up as an unresolvable conflict between the two most general and irreconcilable attitudes about the nature of organic form.

Geoffroy, a romantic and a dreamer, had a vision--perhaps the boldest, the most noble, the most comprehensive idea ever promoted in biology. As a pure morphologist, he sought nature's order in a single abstract plan for all organic design. He maintained, virtually as a motto, "there is, philosophically speaking, only a single animal." (Cubber, as a functional morphologist, denied the utility or reality of such abstractions and argued that an irreducible diversity of form matches the particular functions that organisms perform).

Goethe hoped against hope that Geoffroy could beat the silver-tongued Cuvier, one of France's greatest orators. For Goethe shared Geoffroy's vision and had, himself, advanced for plants the same general argument that Geoffroy now advocated for animals.

Goethe's idealistic morphology of plants, proposed late in the 18th century, tried to understand the varied parts, particularly the diverse structures of flowers, by conceiving them as so many expressions of one underlying, abstract theme that could generate all existing particulars

through a set of geometrical transformations. For Goethe, this ideal source was the leaf itself.

Geoffroy applied the same vision to animals, arguing that vertebrae could perform for moving things the same role that Goethe ascribed to leaves as the foundation of form in plants. The dream that an apparently overwhelming diversity might be reduced to one underlying generality--a basis sought not in the mechanical terms of physics and chemistry, but in the ideal realm of pure form--motivated both men.

But scientific visions, no matter how noble, are only so good as their evidence--and Geoffroy proposed an elaborate set of hypotheses to affirm his claim that animals are built as repeated and modified vertebrae. In particular, he sought to establish a homology--that is, a true similarity based on common origin, rather than a mere analogy, or likeness developed independently by two groups in response to shared functional needs--between the complete segmentation of insects and the partial repetition of vertebrates (seen primarily in the vertebral column and paired ribs).

Such a proposed homology between the segments of an insect's external skeleton and a vertebrate's internal spine lead to some mighty peculiar comparisons, if we pledge to follow the basic model in a forthright and rigorous way--as Geoffroy certainly did. He wrote in 1822, with a firm assurance that others will interpret as bold or foolhardy according to their prejudices:

> I have never ceased to follow nature's great thought: the unity of organic composition; and, in the confidence that this fundamental truth has given me, I must grasp my problem in its greatest generality, approaching each part not with hesitation, but moving on the contrary with sureness.

Two comparisons, in particular, struck most biologists as especially hard to swallow, and helped to seal Geoffroy's fate. First, if the bones of a vertebrate's spine are homologous with external segments of insects, then insects must--quite literally--live within their own vertebrae. Goeffroy wrote in 1822: "every animal lives outside or inside its vertebral column" (tout animal habite en dehors ou en dedans de sa colonne vertebrale). Second, if the

paired attachments to all "vertebrae" are also homologous, then the legs of insects are the same structure as the ribs of vertebrates--and insects walk on their ribs.

The comparison of insect and vertebrate caused difficulty enough for Geoffroy's vision of organic unity., But he wanted more and went further, provoking Cuvier's ire and his own downfall. Early in 1830, two of Geoffroy's young disciples tried to extend the grand homology by arguing that cephalopods (squids and their allies) are built as vertebrates doubled back on themselves--thus bringing the great phylumn of mollusks (including such overtly unsegmented creatures as clams and snails) into line with arthropods and vertebrates.

This was too much for Cuvier, the preeminent functionalist who understood that the strong similarities between squid and fish were analogies independently built as good hydrodynamic solutions to the problems of swimming. The great debate had begun.

Goethe, from the sidelines, rooted for Geoffroy, but most observers scored the match as a decided victory for Cuvier (although debates of such abstraction are not simply "won" or "lost" in conventional terms). And so history has judged this curious event. In our current classifications, complex animals are divided into two major branches, the protostomes (including insects) and the deuterostomes (including vertebrates)--and never the twain shall meet. As a graduate student in systematics 20 years ago, I learned no truth with greater assurance than the proposition, proved by Cuvier against Geoffroy, that the full and complete segmentation of arthropods has no homology with the limited and imperfect segmentation of vertebrates, confined primarily to muscles of the trunk and their associated skeletal and nervous systems. Goeffroy's vision of unity in organic design, died 150 years ago.

But big ideas developed by great thinkers rarely disappear completely. Few men of vision are so profoundly wrong that the motivating dream of an intellectual lifetime contains nothing worthy of later attention, often in an intriguingly altered form. Sure enough, last year's biggest news in developmental genetics has exhumed Geoffroy's old vision of homology between the segmentation of arthropods and vertebrates (although I have been astonished that no

one, in scores of primary articles and secondary reports devoted to this subject of the so-called homeobox, has drawn the obvious connection with Geoffroy's ancient vision). The story, as we shall see, is particularly interesting because a superficial interpretation of this "homeobox" revives Geoffroy's dream in its overt form. We shall see, however, that this interpretation cannot stand--but that victory may yet be snatched from these jaws of Geoffroy's second defeat to affirm something deeper about his curious, yet profound, vision.

In the fruit fly Drosophila, best known of all complex animals in genetic terms, differentiation of segments (why some develop mouth parts, other wings and legs, still others abdominal structures) falls under the control of two gene complexes located on the right arm of the third chromosome. Each complex contains several genes. The antennapedia complex (called ANT-C for short) influences anterior segments of the head and thorax, while the bithorax complex (BX-C) works upon the fly's rear end, from the back part of the thorax through the numerous abdominal segments. Well-known mutations in several of these genes have the peculiar property of placing normal structures in the wrong places. (In the namesakes of the two complexes, flies with the antennapedia mutant grow legs where their antennae should be, while bithorax individuals develop a second pair of wings in opposition to their own name--flies are Diptera, or two-winged). Such mutations, replacing expected organs with homologous parts that should develop in other places are called homeotic--a term coined early in the century by William Bateson, who also gave genetics its name. (See my column of Oct., 1980 for more information on homeotic mutations).

Early in 1984, molecular geneticists announced a remarkable discovery about the fine structure of three genes in these homeotic complexes. We are living in the midst of a revolution in technique within genetics (spilling into public attention via debates about genetic engineering). The methods of research in so-called recombinant DNA have permitted us to learn, in a routine way (albeit with a vast amount of hard work), the sequence of base pairs that builds the genetic material of DNA. We can, in other words, now read the detailed structure of genes.

A very short region, only 180 base-pairs long and

located within the antennapedia gene, showed such remarkable similarity to a unit of identical length in another gene of the antennapedia complex (called fushi tarazu, or "futz" for short among the cognoscenti) that both must share a common evolutionary origin--the similarities are too complex and detailed for independent evolution under similar constraints of function (see my column of April, 1985 for an explicit discussion of when and why detailed molecular similarity must record homology, or unity of evolutionary origin). Soon afterward, the same short sequence was detected in a gene of the other homeotic complex, ultrabithorax of BX-C.

This short, highly conserved sequence received the name "homeobox" because it was first found only in genes of the homeotic complexes and because its trim, tight composition of but 180 base pairs suggested a discrete unit (a "cassette" as one discoverer wrote) that might be copied precisely and inserted into various parts of an organism's genetic material.

This discovery provoked a search for other copies of the homeobox within Drosophila genes. The work continues, but a recent report (a review by W.J. Gehring in Cell, January, 1985, pages 3-5) mentions more than 10 additional genes contining homeoboxes. The homeotic complexes have emerged as preferred "homes" for this conserved sequence--five genes of ANT-C, and three within BX-C, contain homeoboxes. The copies found outside the homeotic complexes lie in genes that also seem to code for some aspects of spatial organization in the embryonic segments.

The homeobox first inspired excitement because it seemed to confirm a longstanding hypothesis about the evolutionary origin of the homeotic complexes. Ed Lewis of CalTech, the world's expert on the homeotic genes of Drosophila, had proposed that the many genes of BX-C (at least eight), arrayed in linear sequence along the right arm of the third chromosome, arose as repetitions of a single ancestral gene (see my column of November, 1981 on the ability of some genes to make copies of themselves and move to new positions). The presence of such a highly conserved, shared sequence within so many homeotic genes provides strong support for Lewis' theory (an alternative, or complementary, idea might hold that only the homeobox "cassette" duplicates and inserts itself in new places).

The simple existence of such a highly conserved sequence in so many places provided enough excitement, but the story goes deeper--thus forging the link to Geoffroy's old vision. (I should say at the outset of this linear account of events that I am not describing discoveries in their chronological order, but in a logical sequence of my own construction that allows me to make sense of much fine work, proceeding in several laboratories and cascading simultaneously in many directions. All these discoveries occurred (or were published) during 1984--a good sign of how quickly an exciting field can move.

The presence of so many copies throughout the Drosophila genome does not prove that the homeobox does anything important for the organism (some repeated elements may act as "selfish DNA," able to replicate, move and insert, but performing no function for the larger organism containing them--again, see my column of November, 1981). An argument for function must first determine whether or not the homeobox gets transcribed into RNA and then translated to protein or enzyme--and the link to Geoffroy's vision requires a hypothesis of function. Much DNA is "silent," or untranscribed, and much may be "junk" (so-called) with reference to its utility for organisms.

The homeobox is, in fact, transcribed and translated--so this basic argument for function holds. Since DNA is a triplet code, the 180 base-pairs of the homeobox produce a so-called "homeodomain" of 60 amino acids. The strongest argument for important function rests upon the discovery that this homeodomain (the 60 amino acids) is even more highly conserved than the parent homeobox (the 180 base pairs of DNA). The foregoing statement may seem puzzling--that higher amino acid than code similarity indicates functional meaning for a stretch of DNA--but the justification arises from the nature of the DNA code.

The genetic code is, in technical parlance, "redundant" at the third position in most cases--that is, the first two base pairs of any triplet determine the amino acid that will be translated, while the third position can be occupied by any of four possible base pairs and still yield the same amino acid in translation. Thus, any mutation at the first or second position will cause the translation of an altered amino acid, but third position mutations are usually "silent" because they induce no change in the product--the

amino acid --that the organism sees and uses.

Now, the homeodomain is substantially more similar among its several copies than the homeobox that codes for it (for the first copies discovered in Drosophila, homeobox similarities ranged between 75 and 79 percent of base pairs, while homeodomain conservation rose to 87 percent of amino acids). The basis of this difference must be--and indeed is, as we now know from direct sequencing of the base pairs and amino acids--that many silent mutations accumulate among the different copies of the homeobox to drive their overall similarity down, but that something is tightly conserving the amino-acid sequences by rejecting first and second position mutations that change the structure and function of proteins, but permitting the "invisible" silent mutations to accumulate in third positions. That "something" must be natural selection, operating to conserve the protein structure of the homeodomain. If natural selection is working so assiduously to conserve the homeodomain, then its protein must be doing something right for the organism--it must, in other words, perform an important function. (The comparison of silent versus translated mutations is a standard test for function in molecular genetics, not a new argument developed for the homeobox story.)

Thus, we know that the homeobox occurs in several Drosophila genes, mostly in the homeotic complexes, and that it codes for a protein with important functions. But what does this protein do? At first glance, but please remember that "first glances" can often be dangerously simplistic, we might be tempted to conjecture that the homeobox must be a key to the control and expression of segmentation. Why else should its copies lie in genes of the homeotic complexes and rarely elsewhere?

We might gain further clues about function of the homeobox by scanning other organisms to see where it does (and does not) occur. Initial results on other invertebrates might be viewed as enhancing the suspicion that homeoboxes are regulators of segmentation. The homeobox sequence has been found in other species of Drosophila, in Tenebrio molitor (a beetle) and in an earthworm of the genus Lumbricius--all segmented animals. It has not been detected in a nematode, a sea urchin and the bacterium E. coli, all unsegmented creatures (see W. McGinnis and others, Cell, volume 37, 1984, pages 403-408).

But the biggest surprise, and the source of popular appeal for the homeobox, arrived with its discovery in yet another group of organisms--our own. If Drosophila is the standard invertebrate of laboratories throughout the world, the South American toad Xenopus has been granted this dubious distinction among cold-blooded vertebrates (mice play the part for mammals). Inevitably, geneticists turned their search for homeoboxes upon Xenopus--with success. Homeoboxes have been found, so far, in two Xenopus genes. (Our knowledge of vertebrate development is so poor that we do not know, for certain, how any single gene operates to regulate the position or differentiation of structures in embryology. But we have some pitifully preliminary hints that the homeobox-containing genes of Xenopus may play some role in development. One gene, for example, produces RNA transcripts in egg cells and again at the crucial stage of gastrulation in early embryology--where the transcripts might be translated to proteins involved in the determination of embryonic cell types, see Muller and others, Cell, volume 39, 1984, pages 157-162).

The similarity between these Xenopus homeoboxes and those discovered earlier in Drosophila is startling. The corresponding homeodomain of one Xenopus gene is identical to the homeodomain of antennapedia in Drosophila at 59 of 60 amino acid positions! This frog-fly concordance is far stronger than the similarity between the two frog homeodomains (54 of 60 amino acids, with 23 silent third-position mutations in addition), or between any pair of homeodomains within Drosophila. Something is working awfully hard to conserve this similarity across one of the most profound phylogenetic gaps in the history of life (the common ancestry of insect and vertebrate predates the establishment of our adequate fossil record in the Cambrian explosion nearly 600 million years ago).

Further scrutiny then revealed homeoboxes in other vertebrate genes. A gene on mouse chromosome 6 contains a homeobox with 66-70 percent base pair similarity to Drosophila and 66 percent to Xenopus (W. McGinnis and others, Cell, volume 38, 1984, pages 75-80). Two human genes also sport homeoboxes, yielding an average of 90 percent amino acid similarity with the three homeodomains first discovered in Drosophila (see Levine and others, Cell, volume 38, 1984, pages 667-673).

As discovery cascaded upon discovery, many scientists and commentators asked what common theme lay behind these obviously interrelated observations. An intriguing possibility inevitably bubbled forth: homeoboxes are primarily found in genes that regulate the differentiation of segments in Drosophila; they are also present in other segmented invertebrates, and absent from unsegmented creatures. Vertebrates also have homeoboxes and at least some form of imperfect segmentation. Of course, we had all learned that segmentation in insects and vertebrates can only be analogous--a product of independent evolution in lineages long separate. Or must segmentation be only analogous? Could it possibly be homologous after all. The homeoboxes of insects and vertebrates are clearly homologous. If they regulate segmentation...well, by God, could Geoffroy have been right all along and after all (following 150 years of calumny and the kind of dismissal that so erased his name from the history of science that not a single geneticist remembered this link with the past while so many discussed the potential homology of segmentation)?

Press reports poured forth. Some began to speak, in overblown terms, about the Rosetta stone of development, biology's greatest mystery. Scientists tended to be more cautious, but several floated the heretical idea of homology for the segmentation of insects and vertebrates--Geoffroy's common ground plan of animal development. W. McGinnis and colleagues wrote, for example (Cell, volume 38, 1984, p. 679):

> If the conserved homeodomain in fruit flies, frogs, mice and humans is involved in the control of segmental development, then it is possible that the segmentally organized animals in both the protostome and deuterostome classes had a common ancestor, and that the metameric body plan has evolved only once in metazoa. (Metamerism refers to the complete segmentation of arthropods; metazoa are the multicellular animals.)

This explicit vindication of Geoffroy remains a possible interpretation of the homeobox, but proper caution has begun to replace overextended enthusiasm, and other less flashy (but in many ways more intriguing)--and (I suspect) more probable--explanations have now been offered.

The superficial interpretation--if it's in homeotic genes and homeotic genes regulate segments--is logically flawed, in fact one of the classically false syllogisms of basic courses. We might as well argue that sneakers produce a high field goal percentage because all great basketball players wear them.

Sure, the homeobox performs an important function, and (from the genes it chooses as bedfellows) it probably has something to do with the embryology of spatial organization. But why should that "something" be the actual control of segmentation itself (the only argument that would affirm Geoffroy's vision in its explicit form)? Think of the more basic activities that genes must superintend before we can even talk about such overt morphological phenomena as segmentation (just as basketball players have to cover their feet with something before they can display any fancy artistry with drives and jump shots)--particularly the production of proteins in the right places, at the right times and in the right amounts. Perhaps the homeobox works for one of these deeply basic functions acting as a vital prerequisite for anything so much later (in a causal chain) and so much more specific as segmentation.

My friends Elizabeth C. and Rudolf A. Raff (Nature, volume 313, 1985, page 185) have suggested such a more general and more "mundane" (their words) function for the homeobox. Translation of RNA to proteins occurs in the cytoplasm. But proteins that will act as basic regulators of development by turning other genes on and off must get back into the nucleus (where the genes reside on their chromosomes). Perhaps, they suggest, the homeodomain simply acts as a "nuclear tag" attaching to other regulating proteins that must enter the nucleus and, perhaps, facilitating their entry. I do not know whether the Raffs' hypothesis has any validity, but it represents the kind of proposal more likely to prove useful than the superficially attractive idea that homeoboxes build segmentation directly.

In fact, we have some hints of evidence for more basic function. Homeoboxes have also been found in two of the so-called mating-type genes of yeast--genes that seem to control the stable determination of cell types in these organisms. Now yeast, obviously, are not segmented. (This argument, however, does not preclude the notion that homeoboxes

regulate segmentation in insects and vertebrates. After all, the homeobox didn't arise mysteriously from nothing. It must have developed from some precursor in an ancestral unsegmented organism. This ancestral proto-homeobox may have worked in one way for ancestors and then been co-opted as a regulator of segmentation in descendants. I discussed the important evolutionary principle of such quirky shifts in function, known to biologists by the awful and senseless name "preadaptation," in last month's column.)

We may, however, read the discovery of homeoboxes in yeast in a more positive light--not as a problem to finesse in preserving the hypothesis that homeoboxes regulate segmenntation, but as overt support for the conjecture that homeoboxes perform a more basic function common to yeast and segmented organisms--some primary control in the differentiation of cell types, for example.

An intriguing hint in this positive direction comes from the apparent homology between proteins coded by the yeast mating genes and proteins in bacteria that may perform the the most basic regulatory activity of all--binding to DNA. (If (If some genes work by regulating the turning on and off of others, then their products--proteins--must be able to bind with the DNA of other genes). Could the homeobox code for such a central protein function as DNA binding? Such a utility would represent function at the most basic level of organic construction--a role far more general than the production of segments. By the time segments form in insects and (imperfectly) in vertebrates, most of what really matters in the regulation of development has already occurred.

So where does this claim for function more general than segmentation leave Geoffroy's dream of a common ground plan for all organisms? Just when his most radical notion of homology between segmentation of insects and vertebrates seemed on the verge of revival in molecular form, along come all the nay-sayers and doubting Thomases of biological caution to point out that presence of a homeobox in a homeotic gene doesn't mean function <u>for</u> segmentation. These modern Cuviers, it seems, have again crushed the dream of unity after such a short and fitful flicker. Or have they?

What was Geoffroy really after? He had a vision of unity and he expressed it in the deepest and most general

terms that his century knew--the generating rules of overt morphology. But Geoffroy dreamed of something deeper than external form--some underlying basis that would produce similarities of form as a necessary and superficial expression of its work. Indeed, though he invoked form faute de mieux, Geoffroy always sought a deeper unity that would generate form as a product. After comparing the segments of a cow and a crab, he once wrote in 1807: "I have always insisted upon an identity which has extended to the least important relation of all, that of form." And, towards the end of his life, following his rout by Cuvier, he wrote in 1835:

> I have, moreover, regenerated this principle (unity of plan and composition), and obtained for it universality of application, by showing that it is not always the organs as a whole, but merely the materials composing each organ that can be reduced to unity.

In other words, Geoffroy dreamed of a unity deeper than the homology of overt segmentation (which he viewed only as the best empirical evidence he could find--the closest approach that techniques of his time would permit to the underlying basis and mystery of development). Somewhere, wherever good biologists go, I think that Geoffroy's spirit is now rejoicing that the homeobox may unite insects and vertebrates in a developmental unity more basic than their final forms--and that the most general rules regulating the conversion of DNA code to adult organism may be common to all creatures.

Surely, in some sense, Geoffroy was correct. Segmentation in insects and vertebrates probably is only analogous, but all organisms have a common evolutionary origin on this planet, and all must therefore share some basic rules of construction inherited as consequences of this history. There must be "unity of plan and composition" at some higher level that transcends the simple and obvious identity of chemical building blocks.

And perhaps Goethe was right after all to ignore the tumultous events in the streets of Paris and to argue instead that an abstract debate in the French Academy of Sciences had far more ultimate importance in human history.

The art of finding timeless essences in apparent trifles, and ignoring contemporary, engulfing tumults as ephemeral, marks the kind of perception that we call genius. Umberto Eco has captured this vital theme as the centerpiece of his wonderful mystery novel, The Name of the Rose. William of Baskerville, that medieval Sherlock, is surrounded by the portentious events of his century, particularly the dispute between papacies of Rome and Avignon. But he solves the Abbey murders because only he (and his antagonist) understnd that these contemporary doings, however enveloping, will pass and be forgotten in time--and that the only important object in the long run, however invisible to all others amidst such tumult, is the library's copy, the only in existence, of a potentially anti-Christian work of Aristotle. Goethe had his priorities straight as well--and the plea of his dying worlds rings through the ages: Mehr Licht (more light).